Title:

Causal inference in psychopathology: A systematic review of Mendelian randomisation studies aiming to identify environmental risk factors for psychopathology

Jean-Baptiste Pingault*[1,2], PhD; Charlotte Cecil[3], PhD; Joseph Murray[4], PhD; Marcus R Munafò[5, 6], PhD; Essi Viding[1], PhD

[1]Division of Psychology and Language Sciences, University College London, London, United Kingdom

[2]MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neurosciences, King's College London, London, United Kingdom

[3]Department of Psychology, Institute of Psychiatry, Psychology and Neurosciences, King's College London, London, United Kingdom

[4]Department of Psychiatry, University of Cambridge, Cambridge, UK

[5]MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, United Kingdom

[6]UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, United Kingdom.

Corresponding author: Dr. Pingault (j.pingault@ucl.ac.uk), Department of Clinical, Health and Educational Psychology, 26 Bedford Way, WC1H 0DS, London United Kingdom.

Funding/Support: JBP is supported by a European Commission Marie Curie Intra-European Fellowship [N° 330699]. JM is supported by a Wellcome Trust fellowship [089963/Z/09/Z]. MRM is a member of the United Kingdom Centre for Tobacco Control Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged. EV is a Royal Society Wolfson Research Merit Award holder.

Abstract

Psychopathology represents a leading cause of disability worldwide. Effective interventions need to target risk factors that are causally related to psychopathology. In order to distinguish between causal and spurious risk factors, it is critical to account for environmental and genetic confounding. Mendelian randomisation studies use genetic variants that are independent from environmental and genetic confounders in order to strengthen causal inference. We conducted a systematic review of studies (N = 19) using Mendelian randomisation to examine the causal role of putative risk factors for psychopathology-related outcomes including depression, anxiety, psychological distress, schizophrenia, substance abuse/antisocial behaviour, and smoking initiation. The most commonly examined risk factors in the reviewed Mendelian randomisation studies were smoking, alcohol use and body mass index. In most cases, risk factors were strongly associated with psychopathology-related outcomes in conventional analyses but Mendelian randomisation indicated that these associations were unlikely to be causal. However, Mendelian randomisation analyses showed that both smoking and homocysteine plasma levels may be causally linked with schizophrenia. We discuss possible reasons for these diverging results between conventional and Mendelian randomisation analyses and outline future directions for progressing research in ways that maximise the potential for identifying targets for intervention.

Keywords: psychopathology, risk factors, Mendelian randomization, causality, smoking, alcohol, body mass index, depression, anxiety, schizophrenia

Contents

Table of Contents

Introduction

Psychopathology is a major public health concern and the leading contributor for years lived with disability world-wide (Whiteford et al., 2013). Identifying modifiable environmental exposures that are causally related to psychopathology is crucial for designing effective evidence-based social policies and interventions. To date, however, establishing whether a correlation between an environmental exposure (e.g. smoking) and a psychopathology-related outcome (e.g. depression) reflects a truly causal relationship or a spurious association remains a major challenge. To address this issue, Mendelian Randomisation (MR) has been developed as a method using genetic information for strengthening causal inference about environmental risk factors in observational research (Davey Smith, 2010). The present systematic review examines MR studies that included psychopathology-related outcomes. In the introduction, we briefly elaborate on causal inference in psychopathology before explaining the principles, strengths and limitations of MR.

Causal inference in psychopathology

A fundamental objective of epidemiological approaches to psychopathology is to identify modifiable environmental exposures that can be targeted in effective interventions. To this end, establishing causality is crucial: an intervention is extremely unlikely to succeed if the targeted environmental exposure is not causally related to the outcome, simply because if exposure A does not cause outcome B, then modifying A will not change B.

Relying on observational studies to identify such potential targets for intervention is problematic. Observational studies suffer from limitations that prevent causal inference such as environmental and genetic confounding (Richmond, Al-Amin, Davey Smith, & Relton,

2014). For instance, smoking and depression are strongly associated in the general population: heavy smokers are more likely than moderate or non-smokers to suffer from depression (Bjørngaard et al., 2013; Lewis et al., 2011). However, many confounding factors could drive this association. For instance, a third variable, such as socioeconomic adversity, which associates with both higher rates of depression (Melchior et al., 2013) and smoking (Pingault et al., 2013) could account for the association. An association between smoking and depression could also arise from partially overlapping genetic risk factors (Kendler et al., 1993). If environmental and/or genetic confounding totally account for the observed association, the relationship between smoking and depression cannot be described as causal.

Considerable effort has been directed at strengthening causal inferences in observational studies (Imai, King, & Stuart, 2008; Jaffee, Strait, & Odgers, 2012; Rutter, 2007; Rutter, Pickles, Murray, & Eaves, 2001), for instance by using statistical innovation such as matching techniques to balance confounders between exposed or non-exposed groups (Ho, Imai, King, & Stuart, 2007; Stuart, 2010; Stuart & Green, 2008), or using cross-cohort comparison to better account for confounding (Lewis, Relton, Zammit, & Davey Smith, 2013). In addition to these techniques, genetically informative designs have addressed the issue of genetic confounding. Family-based designs such as the discordant monozygotic twin or the in vitro fertilization designs account for genetic confounding using known *genetic similarities between family members*. Reviewing the many family-based designs is beyond the scope of this article and can be found elsewhere (D'Onofrio, Class, Lahey, & Larsson, 2014; Jaffee et al., 2012; Lewis et al., 2013). MR is a recently developed method also using genetic epidemiology to strengthen causal inference (Davey Smith & Ebrahim, 2005). However, unlike family-based designs, MR does not require family-based samples and uses direct genotyping instead of relying on *genetic similarities between family members*.

MR and other techniques for strengthening causal inference in observational research are essential for several reasons. First, experimental manipulations of presumed environmental risk factors to test whether they play a causal role are not always possible for ethical reasons (e.g. neglect or maternal smoking). Second, even when feasible, experimental manipulations such as randomised controlled trials suffer from their own limitations (e.g. highly selected volunteers, which may not be representative of the target population, see Imai et al., 2008; Jaffee & Price, 2012). In such cases, MR and other techniques offer unique strengths to test putative causal relations. Finally, many costly randomised controlled trials have failed to confirm associations reported in observational studies (Davey Smith & Hemani, 2014). Therefore, implementing techniques to strengthen causal inference in observational studies is an important step for guiding the choice of appropriate intervention targets before investing in a randomised controlled trial.

Modifiable environmental exposures and genetic instruments

The aim of MR is to test whether a potentially modifiable environmental exposure is causally related to an outcome of interest. The "potentially modifiable environmental exposure" can either be a modifiable behaviour such as smoking or an intermediate phenotype such as cholesterol levels (Sheehan, Didelez, Burton, & Tobin, 2008). MR first involves the identification of a genetic variant that influences the exposure variable. The genetic variant must either alter the level of the exposure itself – e.g. increase in smoking – or mirror its biological effect – e.g. modify cholesterol levels (Bennett, 2010; Davey Smith et al., 2005; Gage, Davey Smith, Zammit, Hickman, & Munafò, 2013). Importantly, the focus of MR is not the genetic variant itself but rather the potential causal relationship between the environmental exposure and the outcome. As such, MR does not aim to identify genetic factors in order to target individuals on the basis of their genotype. Rather, it uses genetic

variants that are known to affect exposure to environmental risk, in order to test whether specific environmental exposures are causally related to the outcomes of interest (Sheehan et al., 2008). These genetic variants are called 'genetic instruments' because they are used in an instrumental fashion – as a mean to an end – to examine causality.

Principles of Mendelian randomisation

MR is based on the following logical proposition: if an exposure A causes an outcome B, then any variable that influences A will also influence B (Davey Smith & Hemani, 2014). The genetic instrument is chosen because it is known to influence A. If there is a causal relationship between A and B, then the genetic instrument that influences A should also influence B. If we observe that the genetic instrument indeed influences B, we can conclude that there is a causal relationship between A and B, provided all assumptions are met (see Figure 1). A critical assumption is that all the effect of the genetic instrument on the outcome B must happen through the exposure A (i.e. similar to full mediation). Provided this assumption is met, if the path from the genetic instrument to A is significant, and if the path from the genetic instrument to B is also significant, then we have all the information needed to estimate the relationship of interest, i.e. the path between exposure A and outcome B.

Once the causal influence between A and B is established, it also follows from the same initial proposition that any other factor influencing A should influence B (additional considerations on this topic can be found in Burgess, Butterworth, Malarstig, & Thompson, 2012). In particular, any intervention that would be successful in modifying the exposure A, should also be successful at modifying the outcome B. For instance, if MR provides evidence that a modifiable behaviour such as smoking is causally related to an outcome such as depression, then an efficient smoking cessation intervention should also have a positive

impact on depression. Similarly, if MR provides evidence that an intermediate phenotype such as high BMI causes cardiovascular diseases, then a diet that successfully reduces BMI should also reduce cardiovascular diseases. Here, the notion of "intermediate phenotype" refers to any phenotype that lies on the causal pathway leading to the outcome of interest. Intermediate phenotypes can therefore be influenced by genes (e.g. by a genetic instrument) or by the environment (e.g. by an intervention such as a diet).

MR is an example of "instrumental variable", an approach developed in economics to better account for confounding in observational research (Gage et al., 2013).

An instrumental variable must meet the following assumptions represented in Figure 1 to enable adequate causal inference (Gage et al., 2013; Sheehan et al., 2008):

I) It must be associated with the exposure;

II) It cannot be associated with the outcome of interest, except via its effect on the exposure. In other words, all the effect of the instrumental variable should be mediated by the exposure (i.e. no direct effect remains and no effect is mediated via another exposure);

III) It must be independent of all variables (measured or not) that confound the relationship between exposure and outcome; and

IV) It must not introduce new confounders to the relationship.

As can be seen, the instrumental variable approach behind MR involves stringent assumptions. For instance, most measures of environmental risk are unlikely to fulfil these assumptions as they tend to cluster together and, consequently, cannot be independent of measured or unmeasured confounders (Davey Smith et al., 2007). Genetic variants retain critical advantages as instrumental variables. Based on Mendel's first law of segregation, alleles segregate at conception independently of the environment. Following Mendel's second

law of independent assortment, genetic variants are inherited independently of each other. Therefore, a genetic variant will generally not be associated with environmental or genetic confounding factors that can bias observational studies (Brion, Benyamin, Visscher, & Davey Smith, 2014; Davey Smith, 2010).

We can use the association between smoking heaviness and depression as a brief concrete example of MR analysis. A genetic variant, the single nucleotide polymorphism (SNP) rs1051730 located in the nicotine acetylcholine receptor gene cluster (CHRNA5-A3-B4), has been used as a genetic instrument to examine this association. This SNP has been repeatedly associated with smoking, the T allele being associated with increased smoking. The first condition for MR analysis is therefore satisfied as the genetic instrument must be robustly associated with the exposure (Bjørngaard et al., 2013; Lewis et al., 2011; Taylor, Fluharty, et al., 2014). The final step is to test whether this genetic instrument is significantly associated with the outcome. Because the genetic instrument is supposed to have an effect on the outcome only via the exposure (see assumption II), this means that, in the absence of the exposure (i.e. among non-smokers), the genetic instrument should have no effect on the outcome. We therefore need to test whether smokers with the T allele are more depressed, while non-smokers with the T allele show no differences in depression status. If this is the case, then we can derive from the instrumental variable approach that increased smoking causally increases the risk of depression. Conversely, if no significant association is found between the genetic instrument and the outcome, this is interpreted as evidence against a causal association. The implementation of MR analysis is straightforward as it relies on a simple test of the association between the genetic instrument and the outcome. More sophisticated statistical techniques are also available, for instance to calculate the size of the causal effect or to deal with several genetic instruments (Davey Smith & Hemani, 2014).

Box1: Analogy between MR and Randomised Controlled Trials (RCT)

The analogy between randomised controlled trials and MR can help to illustrate the principles of the design (see Figure 2). We take the basic example of a trial aiming to assess the role of a smoking cessation intervention (the treatment) on depression (the outcome). The treatment is allocated at random creating two groups, the treatment group and the control group. When comparing outcomes between treatment and control groups, all measured and unmeasured confounders are accounted for by the randomisation – i.e. there should be no significant difference in means of potential confounders between the two groups. As a result, the only possible differences between the two groups are due to the causal effect of the intervention on some modifiable behaviour, here fewer cigarettes per day for instance. In an intent-to-treat analysis, which compares the treatment vs. the control groups independently of compliance to the treatment, the randomly allocated smoking cessation intervention is used as a predictor of depression. Any significant difference in outcome between the two groups is interpreted as reflecting the causal effect of the intervention on the outcome. By corollary, the effect is interpreted as evidence that fewer cigarettes per day lead to less depression, i.e. that the modifiable behaviour and the outcome are causally related. It is important to note, however, that this last causal inference relies on several assumptions, including that the effect of the treatment is mediated only by the exposure. Instead, it is possible that this effect on derpression has nothing to do with less smoking but is due to, for example, an improved social network following the intervention. This is analogous to the second assumption for instrumental variables: the effects of both the intervention (random allocation) and the genetic instrument (‘genetic allocation’) should be fully mediated by the exposure.

Finally, the random treatment allocation in RCTs and the “genetic allocation” in MR both affect the exposure (e.g. smoking), and are used to infer the causal effect of the exposure on the outcome (e.g. smoking on depression). This is because the genetic allocation also creates a situation where carriers versus non-carriers of a given allele only differ on the level of the exposure of interest (e.g. smoking), as environmental and genetic confounders are balanced. Further comments on the analogy between randomised controlled trials and MR can be found elsewhere (Bennett, 2010; Davey Smith & Hemani, 2014; Nitsch et al., 2006).

Limitations of Mendelian randomisation

Here, we briefly review the main limitations of MR (see Davey Smith & Hemani, 2014 for a more extensive review). First, the application of the method is currently limited because: (a) the number of exposure variables for which an adequate genetic instrument is available is small, and (b) large samples are needed to achieve sufficient power to detect causal effects (Brion, Shakhbazov, & Visscher, 2013). Identifying new genetic instruments and increasing

sample sizes will partly mitigate this issue. Second, the assumptions of MR may be violated, which can lead to incorrect causal inference. Main sources of concern are population stratification, linkage disequilibrium and pleiotropy. Population stratification is essentially a problem of the genetic instrument being associated with an ethnicity confounder. When subgroups in the population differ both in disease rates and allele frequencies for the genetic instrument, population stratification becomes a common cause of both and generates a spurious association between the two (Brion et al., 2014; Davey Smith & Hemani, 2014; Gage et al., 2013). Linkage disequilibrium describes instances where genetic variants are correlated with each other more than would be expected by chance (i.e. an exception to the Mendel's law of independent assortment). If the genetic instrument is in linkage disequilibrium with another genetic variant that causes the outcome via another exposure variable, then the causal inference in an MR study may be invalid (Gage et al., 2013; Sheehan et al., 2008). The use of several independent genetic variants to examine the effect of a single exposure can mitigate this issue. Pleiotropy refers to cases where the genetic instrument has an effect on the outcome via multiple intermediate phenotypes, i.e. the second assumption of instrumental variables is violated (although not all types of pleiotropic effects are problematic, see Davey Smith & Hemani, 2014). Several sensitivity analyses can be used to test whether pleiotropy may be an issue. For example, when using MR to study the effects of smoking severity on depression, one can test to confirm that the genetic instrument is associated with the outcome only in smokers and not in non-smokers (Gage et al., 2013). Further comments on pleiotropy can be found in the discussion.

We now review studies that have applied Mendelian randomisation analysis to assess the effects of exposures on psychopathology-related outcomes. We expect to provide insights

regarding the usefulness and limits of MR in psychopathology research as well as to delineate potential research avenues.

Method

Inclusion criteria

The present systematic review aimed to include all studies that applied MR to one or several psychopathology-related outcomes. Given that the MR method was developed relatively recently, all types of psychopathology-related outcomes were considered (e.g. schizophrenia, depression, psychological distress, addictions). In addition, no restriction was applied regarding the operationalisation of psychopathology constructs (i.e. diagnosis or symptoms) and the type of sample (e.g. clinical or population sample).

Search strategy

PubMed and PsycINFO were searched for MR studies. In PubMed, the pre-existing key word "Mendelian Randomisation Analysis"[Mesh] was used as well as the free search – "Mendelian randomization" or "Mendelian randomisation" – to account for American versus British spelling. In PsycINFO, no pre-existing key word was found so the same free search was conducted. Only studies published in English were considered. The last search was performed on the 31st of July 2015.

Study selection

The search in the two databases yielded 838 records, with 591 remaining after filtering for duplicates. Titles and abstracts were then screened and a vast majority of these 591 reports were excluded as they used MR but did not include psychopathology-related outcomes.

Empirical studies not using MR and non-empirical reports (e.g. reviews or comments) were also excluded. If in doubt, the records were retained for the next step. A total of 21 records were selected at this stage. The 21 full-text articles were then assessed for eligibility and 3 were removed (one editorial comment, and two studies that did not directly assess genetic variants). In addition, 1 additional relevant study (Almeida et al., 2009) was found in the references of an included article. The systematic review therefore included 19 studies.

Results

Table 1 presents a synopsis of studies included in the systematic review. Among the 19 studies included, 14 examined one or a combination of outcomes related to anxiety, depression and psychological distress. We first examine these studies before turning to the others, which were on schizophrenia, substance use/antisocial personality disorder and smoking initiation.

Studies on anxiety, depression and psychological distress

The exposure variables examined in these 14 studies were smoking (n = 4), alcohol intake (n = 2), body mass index (n = 5), fatty acids (n = 1) and C-reactive protein (n = 2).

Smoking. The single nucleotide polymorphism (SNP) located in the nicotine acetylcholine receptor gene cluster (CHRNA5-A3-B4) was used to study effects of smoking in four MR studies (Bjørngaard et al., 2013; Lewis et al., 2011; Taylor, Fluharty, et al., 2014; Wium-Andersen, Ørsted, & Nordestgaard, 2015). It is worth noting that not all these studies are independent as two (Bjørngaard et al., 2013; Lewis et al., 2011) were included in the consortium analysis (Taylor, Fluharty, et al., 2014). As expected, all four studies using this SNP found that that the risk allele was associated with increased smoking heaviness as measured by the number of cigarettes smoked. All four studies also reported that increased

smoking was substantially associated with higher levels of anxiety, depression or psychological distress in conventional analyses adjusting for covariates. However, MR analyses showed that the risk allele did not predict higher levels of depression, anxiety or psychological distress in smokers. As laid out in the introduction, the absence of a significant relationship between the genetic instrument (that predicts environmental exposure) and the outcome is interpreted as evidence against the existence of a causal association. Therefore, the findings of these studies do not support the notion that smoking heaviness causes any of these outcomes. In one study (Lewis et al., 2011), among women who smoked pre-pregnancy, those with the risk allele smoked more during pregnancy but were less likely to report high levels of depressed mood at 18 weeks of pregnancy. These results are consistent with a self-medication hypothesis, whereby smoking is used to alleviate symptoms of depression. One of these four studies (Wium-Andersen, Ørsted, & Nordestgaard, 2015) included several other psychopathology-related outcomes (i.e. antidepressant medication use, schizophrenia and antipsychotic medication use) and compared the results with those for chronic obstructive pulmonary disease. Results for antidepressant medication use were similar to those found for depression, i.e. no evidence of a causal association. As expected, MR analysis indicated a causal role of smoking on chronic obstructive pulmonary disease. Results for schizophrenia and antipsychotic medication are discussed below together with other studies on schizophrenia.

Alcohol. One study examined the relationship between alcohol use and depression in men aged over 65 years (Almeida, Hankey, Yeap, Golledge, & Flicker, 2014). A variant of the alcohol dehydrogenase 1B (ADH1B) gene, which reduces the ability to oxidize ethanol, was used as a genetic instrument. Faster metabolizers tend to consume more alcohol. The average number of drinks per day was related to depression in standard regression analyses adjusting

for covariates. However, in the MR analyses, the genetic instrument was not associated with depression and therefore provided evidence against causality. The other study (Wium-Andersen, Ørsted, Tolstrup, & Nordestgaard, 2015) examined alcohol use in relation to depression and psychological distress. In analyses with adjustment for covariates, alcohol use was not related with prescription of antidepressants or hospitalization/death with depression but was related with self-reported antidepressant use and items assessing psychological distress. However, the genetic variants were not related to any outcome, thus providing evidence against the notion that alcohol use is causally related to any of these outcomes. Interestingly, this study also included the outcome hospitalization/death with alcoholism. Contrary to what was observed for depression, the MR analysis supported a causal relationship between alcohol use and hospitalization/death with alcoholism.

Body Mass Index (BMI). Five studies used BMI as an intermediate phenotype that can be modified by an environmental intervention. The aim was to assess whether BMI was causally related with various measures of anxiety, depression and psychological distress. Four studies used a variant in the FaT mass and Obesity-associated (FTO) gene (Hung et al., 2014; Kivimäki, Jokela, Hamer, et al., 2011; Lawlor et al., 2011; Walter et al., 2015) and 3 used a polygenic risk score (Hung et al., 2014; Jokela et al., 2012; Walter et al., 2015). All studies found that, in conventional analyses adjusting for covariates, higher BMI was associated with higher levels of their respective measures of anxiety, depression and psychological distress. However, MR findings were somewhat contradictory between the five studies. No evidence for a causal effect of BMI was detected in one study of major depression (Hung et al., 2014) and another using a standardized depression scale (Walter et al., 2015). A third study reported evidence that higher BMI was causally related to more self-reported depression and anxiety symptoms only in men (Kivimäki, Jokela, Hamer, et al., 2011). A fourth study using a

polygenic risk score as a genetic instrument provided evidence that higher BMI increased the risk of depressive symptoms, in particular in adolescents (Jokela et al., 2012). Finally, MR analyses in the largest study suggested that the association was in the opposite direction, with higher BMI being related to less psychological distress (Lawlor et al., 2011). This latter finding supports the "fat-jolly" versus the "fat-sad" hypothesis (Kivimäki, Jokela, & Batty, 2011). These contradictory findings in MR analyses may be explained in various ways. First, both the constructs and the assessment methods varied widely, from a clinical assessment of depression to a 4-item measure of psychological distress. Second, the MR assumption that the effect of the genetic instrument should be entirely mediated by the intermediate phenotype – here BMI – was not always met (Kivimäki, Jokela, Hamer, et al., 2011; Walter et al., 2015), casting doubt on the validity of the analyses. Overall, the MR analyses showed that results from conventional analyses adjusting for covariates, which favour the "fat-sad" hypothesis, should be considered carefully until additional evidence is gathered.

Omega-3. One study (Sallis, Steer, Paternoster, Davey Smith, & Evans, 2014) examined the association between two omega-3 Fatty Acids - docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA) – and depression-related outcomes, including perinatal-onset depression, antenatal depression and postnatal depression. Omega-3 fatty acids might affect brain function and behaviour and observational studies report an inverse relationship between fish consumption (a major source of omega-3) and depression. To test the likelihood of a protective causal effect of fatty acids on depression, the authors used polygenic risk scores predicting DHA and EPA. However, only the polygenic score predicting DHA was significant in the study sample so that no MR analysis using EPA was conducted. Finally, the evidence for an association between fatty acids and depression was weak in conventional analyses adjusting for covariates and null in MR analysis.

C-reactive protein. A study of men aged over 65 years examined whether the inflammatory biomarker C-reactive protein, which rises after tissue injury or bacterial infection, was causally linked to depression, using 2 SNPs in the C-reactive protein (CRP) gene (Almeida et al., 2009). A higher level of C-reactive protein was linked to higher depression in bivariate analyses but this was no longer the case in analyses adjusting for covariates. MR analyses also suggested that this relationship was unlikely to be causal. One allele associated with lower C-reactive protein levels was actually associated with higher depression, suggesting a causal association in the opposite direction. However, a large population-based study did not confirm this finding when examining the association between C-reactive protein, depression, psychological distress and major somatic diseases (Wium-Andersen, Ørsted, & Nordestgaard, 2014a). In this study, conventional analyses showed significant relations between C-reactive protein and all outcomes including hospitalization with depression, anti-depressant use, psychological distress and several somatic diseases (e.g. cancer, ischemic heart disease, all-cause mortality). However, none of these associations were significant in MR analyses and, in some cases, the MR estimates were significantly lower than the conventional estimates, suggesting that conventional analyses are biased by confounding variables even after adjustment for covariates.

Studies of schizophrenia.

C-reactive protein. The relationship between C-reactive protein and schizophrenia was investigated in the same sample as above as it has been suggested that inflammation could be involved in the pathogenesis of schizophrenia (Wium-Andersen, Ørsted, & Nordestgaard, 2014b). In conventional analyses, elevated plasma levels of C-reactive protein were associated with late-onset schizophrenia. The corresponding effect size was similar in MR analyses but it was not significant. Because the effect sizes were similar in the two methods,

the authors concluded that a causal relationship between C-reactive protein and schizophrenia could not be excluded.

Smoking. In addition to depression-related outcomes, Wium-Andersen et al. (2015) also investigated the effects of smoking on two separate schizophrenia-related outcomes based on Danish national registries: schizophrenia (i.e. ever being diagnosed with schizophrenia) and antipsychotic medication use (i.e. ever being prescribed antipsychotic medication). The genetic instrument for smoking was significantly associated with antipsychotic medication use in ever smoker but not in never-smokers, suggesting a causal effect of smoking on antipsychotic medication use (i.e. increased smoking leading to increased medication use). The direction of effect was similar for diagnosed schizophrenia but failed to reach significance. The authors then used summary data from the Psychiatric Genomics Consortium to verify whether the genetic instrument for smoking was significantly predictive of diagnosed schizophrenia, which was indeed the case. However, they were not able to distinguish between smokers and non-smokers in this analysis, which would have provided additional evidence that the effect of the genetic instrument was indeed mediated by smoking. These results tentatively indicate that higher smoking levels might causally lead to a higher risk of schizophrenia, contrary to findings on smoking and depression.

Homocysteine. Homocysteine is an amino acid found in blood plasma that, in excess, is associated with cardiovascular diseases and mental retardation. Two meta-analyses conducted by the same research group investigated the potential causal role of homocysteine in schizophrenia. The second meta-analysis (Numata et al., 2015) is an extension of the first one (Nishi et al., 2014) and included the same and additional studies producing a total of 36 case-control studies. The genetic instrument, Methylenetetrahydrofolate reductase (MTHFR), significantly lowered plasma levels of homocysteine. It was also significantly associated with

schizophrenia risk, suggesting that higher levels of homocysteine are causally related to a higher risk of schizophrenia.

Study of substance use and antisocial behaviour.

A small study of 180 Asian adolescent adoptees tested the gateway hypothesis, i.e. that drinking behaviour in adolescence causes the misuse of other psychoactive substances, and antisocial personality disorder (Irons, McGue, Iacono, & Oetting, 2007). The aldehyde dehydrogenase 2 (ALDH2) gene was used as a genetic instrument in MR analysis: the deficient form of the corresponding enzyme, which is frequent in Asian populations, leads to decreased amount of ingested alcohol (because of unpleasant effects of drinking alcohol among people with this variant, such as facial flushing and nausea). Adolescent drinking behaviours varied significantly in the expected direction as a function of the genetic instrument. However, the genetic instrument was not significantly associated with any of the substance use and antisocial outcomes. Therefore, these findings do not support the hypothesis that alcohol use in adolescence causes the misuse of other psychoactive substances, and antisocial personality disorder.

Offspring smoking initiation.

The relationship between maternal smoking during and after pregnancy and offspring smoking initiation during adolescence (14-16 years) was investigated in a longitudinal study (Taylor, Howe, et al., 2014). A maternal SNP in the nicotine acetylcholine receptor gene cluster (CHRNA5-A3-B4) was used. Mendelian randomisation analyses did not show evidence that the relationship between maternal smoking was causally related to offspring smoking initiation and progression to regular smoking.

Discussion

In this systematic review, we examined studies using the MR design to assess whether a range of potentially modifiable environmental exposures were causally related to psychopathology-related outcomes. Conventional analyses adjusting for covariates showed significant associations between exposures and psychopathology-related outcomes in the expected directions in nearly all studies. However, in most cases, these relations were not supported in MR analyses, suggesting that the observed associations were not causal. Based on MR analyses, there was some evidence for a causal effect of BMI on depression, anxiety and psychological distress but the direction of the findings was not consistent. We first discuss possible reasons for these null and contradictory findings. Encouragingly, recent MR studies suggested a possible causal effect of two exposures – smoking and homocysteine levels – on schizophrenia. We discuss these results together with future directions in MR analyses of psychopathology.

Reasons for null and contradictory findings

"Most research findings are false" (Ioannidis, 2005)

Most evidence for relationships between exposures and psychopathology-related outcomes comes from exploratory epidemiological studies. However, in such observational studies, even those that are well-powered and well conducted, most of the reported significant findings are likely to be false. This stems in part from the large number of associations tested and the threshold commonly used for significance testing (alpha = 0.05). Therefore, it may not be surprising that studies trying to replicate these associations will fail in most cases. Furthermore, even consistent research findings may simply reflect accurate measures of bias

in a given field (Ioannidis, 2005). This is particularly important here because the aim of MR is precisely to remove bias due to confounding (i.e. conventional analyses cannot adjust adequately for confounding and therefore provide biased estimates of the true causal relationship). The fact that most included studies found a significant relationship between exposures and psychopathology in conventional analyses is not very conclusive because: (i) conventional regression techniques do not always reliably remove bias (Stuart, 2010); (ii) included studies controlled only for a limited number of observed confounding variables. A good example can be found in the study of C-reactive protein and depression, psychological distress and major somatic diseases (Wium-Andersen, Ørsted, & Nordestgaard, 2014a): C-reactive protein was significantly related to most outcomes in conventional analyses but unrelated to any outcome in MR analyses. As pointed by the authors, the estimates from conventional and MR analyses were significantly different in several instances. This means that MR estimates were not only non-significant but also significantly smaller than estimates derived from conventional analysis, suggesting that conventional estimates were inflated by confounding. Therefore, because confounding is so widespread, a large proportion of significant associations identified using conventional analyses are not likely to be causal, as reflected in non-significant findings based on MR analyses.

Alternative causal explanations: reverse causation, timing and sub-populations

Although the majority of included MR studies suggested that exposures did not have causal effects on the psychopathology-related outcomes, this does not exclude other possible causal relationships between study variables. Reverse causation is a first possibility. For instance, increased smoking did not seem to cause depression, anxiety or psychological distress. However, it is still possible that depression causes increased smoking (Lewis et al., 2011). Similarly, high levels of C-reactive protein (signalling inflammation) might be a

consequence rather than a cause of a number of diseases (Brion et al., 2014). Second, effects may change in direction over time. For instance, increased BMI may be related to less psychological distress in the short-term given comforting effects of eating, but may lead to more psychological distress on the long-term as adverse physical and social consequences accumulate. Finally, causal associations may be stronger in sub-groups of the population (e.g. obesity and depression in adolescents, see Jokela et al., 2012), or exist only under challenging circumstances (e.g. fatty acids and depression during pregnancy, see Sallis et al., 2014)

Heterogeneity in sample size, measures and genetic instruments

Included studies were heterogeneous at several levels, in particular regarding sample size, quality of measures and genetic instruments. Sample size varied from a single study of 180 participants (Irons et al., 2007) to a consortium analysis of over 127,000 participants (Taylor, Fluharty, et al., 2014). As such, lack of statistical power may explain some of the null findings. Although research regarding power in MR is emerging, some studies were clearly underpowered. For example, with a genetic instrument explaining 1% of the variance in the exposure, an observed correlation of 0.40 and a true causal correlation between the exposure and the outcome of 0.20, almost 20,000 participants are needed to achieve 80% power. Half of the reviewed studies included less than 5,000 participants. However, the larger studies had power to detect even small causal effects. For example, the null findings regarding smoking and depression in very large samples make it unlikely that there is any substantial causal effect (see Brion et al., 2013 for power calculation). In addition, confidence intervals are typically larger in MR analyses compared to conventional analyses, which makes it harder to detect significant effects. For instance, the effect sizes of the association between C-reactive protein and schizophrenia were similar in conventional and MR analysis, but not significant

using MR despite the very large sample size (Wium-Andersen, Ørsted, & Nordestgaard, 2014b).

There was very considerable heterogeneity in how psychopathology-related outcomes were measured, from single items assessing psychological distress to clinical diagnoses of depression. This adds to the notorious difficulty in the phenotypic characterization in psychopathology, as, for instance, patients with very little overlapping symptoms may be given the same diagnosis (Sallis et al., 2014; Wium-Andersen, Ørsted, & Nordestgaard, 2014a). Such heterogeneity does not facilitate the discovery of reliable environmental exposures causally associated with psychopathology. Some heterogeneity was also observed in the genetic variants that were used. The SNP rs1051730 in the nicotine acetylcholine receptor gene cluster (CHRNA5-A3-B4), a gene cluster with a well-documented biological function was used in studies of smoking and depression. Conversely, two different SNPs with unclear biological function and/or polygenic risk scores with more than 30 SNPs were used to assess the consequences of increased BMI (although a recent study shed a light on the function of rs1421085, a SNP in the FTO region used in Kivimäki, Jokela, Hamer, et al., 2011, which seems to have pronounced effects on obesity through adipocyte thermogenesis regulation, see Claussnitzer et al., 2015).

Overall, the use of single variants versus polygenic scores or variants with known versus unclear biological functions affect the power to detect an association as well as whether the assumptions of MR are plausible.

Assumption violations

The possibility must also be considered that results from conventional and MR analyses differed because assumptions of MR analyses were not fulfilled in some studies. As detailed in the introduction, a number of assumptions must be met for MR to yield unbiased estimates

of a causal effect. Included studies often presented the associations between the genetic instruments and potential confounders, which, apart from rare exceptions, were all non-significant. This shows that genetic instruments seemed largely free from bias due to confounding, whilst still strongly predicting the exposure of interest. Therefore, MR analysis did seem to have successfully removed bias due to confounding. However, another assumption is that the effect of the genetic variant on the outcome is entirely mediated by its effect on the exposure. This assumption was not always met. In particular, some associations between the genetic instruments used for BMI and outcomes remained largely unchanged when controlling for BMI (Lawlor et al., 2011). This means that these genetic variants might affect the outcome by pathways other than BMI, violating this assumption and biasing causal inference. Such violations of the assumptions may explain some of the aforementioned contradictory findings regarding the effects of BMI.

Future directions: Increasing the scope of MR studies

At this stage, it is important to note that MR analyses have uncovered many likely causal associations outside the field of psychopathology, for instance between alcohol intake and oesophageal cancer, and tobacco smoking and BMI (Brion et al., 2014). Therefore, the numerous null findings in the present review may simply stem from the fact that, in most cases, there were simply no true causal relations between the exposures examined and the psychopathology-related outcomes under scrutiny. However, three recent reports suggested a potential causal effect of two exposures – i.e. homocysteine and smoking – on schizophrenia. Although replication is warranted, these findings are encouraging with regard to the potential of MR studies to further our understanding of psychopathology. The application of MR to psychopathology is recent. In the next decade, the scope of applications of MR in the field

should increase rapidly with new research questions being asked and new technical and methodological innovations being implemented.

The vast majority of studies included in this review examined outcomes related to depression, anxiety and psychological distress. There seems to be no valid reason why the scope of psychopathology-related outcomes examined in MR studies should remained so restrictive. For instance, replications of the study on substance use and antisocial behaviour (Irons et al., 2007) in larger samples are warranted. Other relevant questions would also benefit from an MR approach, such as further elucidating the nature of the relationship between smoking and Attention Deficit and Hyperactivity Disorder (McClernon & Kollins, 2008; Pingault et al., 2013).

Genome-wide information is increasingly available on large datasets, which will increase the scope of MR in two ways. First, with more data available, it will become easier to have adequate sample size to examine new research questions (see Taylor, Fluharty, et al., 2014 for an example of consortium analysis). MR analysis can often be implemented using only summary data from genome-wide association studies (GWAS), which enables the hypothesis-free examination of relationships between a large number of genetic instruments and outcomes (Evans & Davey Smith, 2015). Second, GWAS are identifying an increasing number of genetic variants associated with psychopathology risk factors and disorders. With these additional instruments, it will be possible to address new questions using MR. For example, discovering a genetic instrument for cannabis use would provide new insights in the long-standing dispute over the putative causal role of cannabis in the onset of schizophrenia.

A promising avenue to generate new genetic instruments is the use of polygenic risk scores or the simultaneous use of multiple SNPs (Burgess & Thompson, 2013; Davies et al., 2015). Using multiple SNPs has several advantages. First, combining SNPs increases the

variance explained by the instrument, which is a critical parameter if we are to achieve sufficient power in MR analysis. Second, the use of multiple SNPs opens new research avenues in case no single SNP instrument is available, which will be particularly useful in the case of psychopathology. Indeed, GWAS on psychopathology phenotypes have clearly shown that finding any common genetic variant with a large effect is unlikely. Instead, many SNPs with small effects have been found for some disorders (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) or are yet to be found for others (Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium et al., 2013).

Particularly promising is the possibility of using multiple SNPs in a two-way MR design to test reciprocal causal influences. For instance, the studies included in this review indicated that heavier smoking does not seem to cause increased depression. However, reverse causation is possible as depression may lead to increased smoking. If GWAS successfully identify multiple genetic variants that are associated with depression and fulfil the criteria for genetic instruments, it then becomes possible to examine this reverse causation hypothesis. Similarly, one study included in this review (Wium-Andersen, Ørsted, & Nordestgaard, 2015) suggested a causal pathway from smoking to schizophrenia. More than a hundred of genome-wide significant SNPs associated with schizophrenia have been uncovered in the latest GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) and can be used to investigate the reciprocal potential causal pathway from schizophrenia to smoking.

Although promising, the use of multiple SNPs has its own drawbacks. A major one is horizontal pleiotropy, which happens when a genetic variant has a direct effect on both the exposure and the outcome. Horizontal pleiotropy violates the second MR assumption (i.e. that all the effect of the genetic instrument on the outcome must be mediated by the exposure). When using multiple SNPs, the likelihood is high that horizontal pleiotropy will be present

for some of these SNPs. This is a real concern as pleiotropy is widely spread throughout the phenome, including for psychopathology-related outcomes (Evans et al., 2013; Krapohl et al., In Press). Accounting for this potential bias induced by the use of multiple SNPs is an active area of research. In particular, a recent method derived from meta-analytic techniques, the Egger regression, can be used to estimate pleiotropic effects and generate a causal estimate corrected for pleiotropy (Bowden, Davey Smith, & Burgess, 2015).

The number of exposures relevant to psychopathology, which are possible to examine using MR, will likely substantially increase due to the discoveries of new single SNP instruments and advanced methods for using multiple SNPs. MR studies using few genetic variants related to intermediate phenotypes through clear biological pathways (e.g. C-reactive protein or uric acid) offer the highest degree of causal evidence (Burgess, Timpson, Ebrahim, & Davey Smith, 2015). These studies can be used to validate or invalidate potential targets for drug development. For instance, drugs targeting C-reactive protein were not further developed by pharmaceutical companies after MR studies showed no evidence of its causal role in cardiovascular diseases (Burgess et al., 2015). This example demonstrates the utility of null MR findings in reducing the number of potential intervention targets, and therefore the costs of drug development. The same should soon be possible with psychopathology-related outcomes. For instance, studies included in this review point towards homocysteine as a relevant target for drug development in schizophrenia. In the case of complex intermediate phenotypes influenced by many genetic variants of small effects (e.g. BMI and depression), additional caution in the interpretation is warranted given the increased likelihood of horizontal pleiotropy and biological pathways that are not well understood. In this case, MR studies can be used to probe the plausibility of a causal relationship and investigate potential causal pathways (Burgess et al., 2015).

Finally, it is worth noting that adequate genetic instruments may never be found for a number of important risk factors for psychopathology, for instance neighbourhood or parenting characteristics, in which case alternative methods to strengthen causal inference (e.g. discordant monozygotic twin design) may be used.

Box 2: MR in the "omics era": understanding developmental mechanisms

The scope of MR will also increase with the availability of "omics" data such as proteomic data or epigenetic data (Brion et al., 2014). Here, we focus on epigenetics as a potential 'missing link' in the aetiology of complex disorders. Epigenetic mechanisms, such as DNA methylation, influence dynamic changes in transcription independently of genomic sequence (Jaenisch & Bird, 2003). Altered DNA methylation patterns have been shown to associate with both environmental risk exposure such as prenatal diet or childhood maltreatment (Lutz & Turecki, 2014; Tobi et al., 2014) as well as a range of psychopathology-related outcomes, for instance major depression, posttraumatic stress and schizophrenia (Fuchikami et al., 2011; Klengel, Pape, Binder, & Mehta, 2014; Wockner et al., 2014). Together, these findings have provided initial support for the role of epigenetic processes as a *mediator* in the link between environmental influences and psychopathology-related outcomes. However, establishing the causality of epigenetic associations has been challenging, not least because epigenetic markers are equally as susceptible to confounding, measurement error and reverse causation as the environmental exposures themselves (Relton & Davey Smith, 2012a). For example, it is presently unclear whether altered DNA methylation patterns are truly a risk factor for – or a consequence of – psychopathology-related outcomes. Consequently, epigenetic epidemiology needs to integrate causal inference methods.

In response to this need, a two-step epigenetic extension to MR has recently been proposed to test causal mediation (Relton & Davey Smith, 2012b). While the method was originally developed with epigenetics in mind, it may be applied to other mediating variables as well (Brion et al., 2014). In two-step MR, the first step assesses the causal association between the independent variable (e.g. modifiable environmental exposure) and the mediator (e.g. DNA methylation), while the second step assesses the causal association between the mediator and the dependent variable (e.g. psychopathology-related outcome). Each step can also be performed in isolation. In Step 1, a genetic variant is used as a proxy for an environmental exposure. As with standard MR, causality is supported only if the genetic proxy is associated with the epigenetic marker *exclusively* through its association with the environmental exposure. With regards to the mediator, several strategies have been proposed for identifying appropriate DNA methylation markers, including the application of epigenome-wide and candidate gene strategies (Relton & Davey Smith, 2012a). In Step 2, a genetic proxy for DNA methylation – specifically, a local genotype (*cis*- SNP) – is used to obtain an unbiased estimate for the effect of DNA methylation on the psychopathology-related outcome. In conjunction, these steps can clarify whether epigenetic processes mediate environmental effects on a psychopathology-related outcome.

Although two-step MR shows promise as a method for testing causal mediation, it is still in its infancy and currently lacks empirical investigation, particularly with regards to psychiatric phenotypes (Kirkbride et al., 2012). Furthermore, epigenetic MR faces a number

of additional challenges on top of those previously described in relation to standard MR (e.g. population stratification, LD and pleiotropy). First, epigenetic associations are usually only modest in size (Brion et al., 2014). Second, because of low statistical power, the MR design requires a much larger sample size than what is typically available for epigenetic studies. This is particularly true for studies that examine methylation in central, as opposed to peripheral (e.g. blood, saliva) tissues. Third, in contrast to genomic sequence which remains fixed, epigenetic markers have been shown to vary both across tissues and across time (Mill & Heijmans, 2013). Such sources of variability may greatly influence the associations under investigations, as epigenetic effects may be limited to specific tissues or developmental periods. Despite the above challenges, the application of two-step MR may offer novel insights into causal developmental pathways, as well as elucidating whether environmental exposures 'get under the skin' to influence psychopathology-related outcomes via epigenetic processes.

Limitations

The main limitation of the present systematic review is that we did not conduct meta-analyses. Meta-analyses are particularly interesting in MR as they increase the power to detect any causal association. Several reports included in this review used meta-analyses (Numata et al., 2015; Taylor, Fluharty, et al., 2014). However, meta-analyses seem premature for most pairs of exposure-outcome variables at this stage as few studies were available and, for each of these pairs, the heterogeneity in outcome definition and genetic instruments was considerable.

Conclusions

We systematically reviewed studies that applied the MR design to psychopathology-related outcomes. In conventional analyses, associations between exposures and outcomes were generally significant and in the expected directions. However, MR analyses often contradicted these results by providing very little consistent evidence that any of these associations were causal. These findings highlight the potentially strong bias in conventional analyses of risk factors in psychopathology. The implications are far reaching in that even replicated findings

based on conventional designs and analyses may not be reliable enough to meaningfully inform drug development as well as preventive interventions and policies. Most recent studies and methodological innovations highlight the potential of well-designed and well-powered MR studies to contribute to a better identification of relevant causal risk factors in psychopathology.

References

Almeida, O. P., Hankey, G. J., Yeap, B. B., Golledge, J., & Flicker, L. (2014). The triangular association of ADH1B genetic polymorphism, alcohol consumption and the risk of depression in older men. *Molecular Psychiatry*, *19*(9), 995–1000. http://doi.org/10.1038/mp.2013.117

Almeida, O. P., Norman, P. E., Allcock, R., van Bockxmeer, F., Hankey, G. J., Jamrozik, K., & Flicker, L. (2009). Polymorphisms of the CRP gene inhibit inflammatory response and increase susceptibility to depression: the Health in Men Study. *International Journal of Epidemiology*, *38*(4), 1049–1059. http://doi.org/10.1093/ije/dyp199

Bennett, D. A. (2010). An introduction to instrumental variables--part 2: Mendelian randomisation. *Neuroepidemiology*, *35*(4), 307–310. http://doi.org/10.1159/000321179

Bjørngaard, J. H., Gunnell, D., Elvestad, M. B., Davey Smith, G., Skorpen, F., Krokan, H., … Romundstad, P. (2013). The causal role of smoking in anxiety and depression: a Mendelian randomization analysis of the HUNT study. *Psychological Medicine*, *43*(4), 711–719. http://doi.org/10.1017/S0033291712001274

Bowden, J., Davey Smith, G., & Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*, *44*(2), 512–525. http://doi.org/10.1093/ije/dyv080

Brion, M.-J. A., Benyamin, B., Visscher, P. M., & Davey Smith, G. (2014). Beyond the single SNP: Emerging developments in Mendelian randomization in the 'omics' era. *Current Epidemiology Reports*, *1*(4), 228–236. http://doi.org/10.1007/s40471-014-0024-2

Brion, M.-J. A., Shakhbazov, K., & Visscher, P. M. (2013). Calculating statistical power in Mendelian randomization studies. *International Journal of Epidemiology*, *42*(5), 1497–1501. http://doi.org/10.1093/ije/dyt179

Burgess, S., Butterworth, A., Malarstig, A., & Thompson, S. G. (2012). Use of Mendelian randomisation to assess potential benefit of clinical intervention. *BMJ (Clinical Research Ed.)*, *345*, e7325.

Burgess, S., & Thompson, S. G. (2013). Use of allele scores as instrumental variables for Mendelian randomization. *International Journal of Epidemiology*, *42*(4), 1134–1144. http://doi.org/10.1093/ije/dyt093

Burgess, S., Timpson, N. J., Ebrahim, S., & Davey Smith, G. (2015). Mendelian randomization: Where are we now and where are we going? *International Journal of Epidemiology*, *44*(2), 379–388. http://doi.org/10.1093/ije/dyv108

Claussnitzer, M., Dankel, S. N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., … Kellis, M. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. New England Journal of Medicine, 373(10), 895–907. http://doi.org/10.1056/NEJMoa1502214

Davey Smith, G. (2010). Mendelian randomization for strengthening causal inference in observational studies: application to gene × environment interactions. *Perspectives on Psychological Science*, *5*(5), 527–545. http://doi.org/10.1177/1745691610383505

Davey Smith, G., & Ebrahim, S. (2005). What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ*, *330*(7499), 1076–1079. http://doi.org/10.1136/bmj.330.7499.1076

Davey Smith, G., Ebrahim, S., Lewis, S., Hansell, A. L., Palmer, L. J., & Burton, P. R. (2005). Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet*, *366*(9495), 1484–1498. http://doi.org/10.1016/S0140-6736(05)67601-5

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, *23*(R1), R89–98. http://doi.org/10.1093/hmg/ddu328

Davey Smith, G., Lawlor, D. A., Harbord, R., Timpson, N., Day, I., & Ebrahim, S. (2007). Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Medicine*, *4*(12), e352. http://doi.org/10.1371/journal.pmed.0040352

Davies, N. M., von Hinke Kessler Scholder, S., Farbmacher, H., Burgess, S., Windmeijer, F., & Smith, G. D. (2015). The many weak instruments problem and Mendelian randomization. *Statistics in Medicine*, *34*(3), 454–468. http://doi.org/10.1002/sim.6358

D'Onofrio, B. M., Class, Q. A., Lahey, B. B., & Larsson, H. (2014). Testing the developmental origins of health and disease hypothesis for psychopathology using family-based quasi-experimental designs. *Child Development Perspectives*, *8*(3), 151–157. http://doi.org/10.1111/cdep.12078

Evans, D. M., Brion, M. J. A., Paternoster, L., Kemp, J. P., McMahon, G., Munafò, M., … Davey Smith, G. (2013). Mining the human phenome using allelic scores that index biological intermediates. *PLoS Genetics*, *9*(10), e1003919. http://doi.org/10.1371/journal.pgen.1003919

Evans, D. M., & Davey Smith, G. (2015). Mendelian randomization: New applications in the coming age of hypothesis-free causality. *Annual Review of Genomics and Human Genetics*. http://doi.org/10.1146/annurev-genom-090314-050016

Fuchikami, M., Morinobu, S., Segawa, M., Okamoto, Y., Yamawaki, S., Ozaki, N., … Terao, T. (2011). DNA methylation profiles of the brain-derived neurotrophic factor (BDNF) gene as a potent

diagnostic biomarker in major depression. *PloS One*, *6*(8), e23881. http://doi.org/10.1371/journal.pone.0023881

Gage, S. H., Davey Smith, G., Zammit, S., Hickman, M., & Munafò, M. R. (2013). Using Mendelian randomisation to infer causality in depression and anxiety research. *Depression and Anxiety*, *30*(12), 1185–1193. http://doi.org/10.1002/da.22150

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199 – 236. http://doi.org/http://dx.doi.org/10.1093/pan/mpl013

Hung, C.-F., Rivera, M., Craddock, N., Owen, M. J., Gill, M., Korszun, A., … McGuffin, P. (2014). Relationship between obesity and the risk of clinically significant depression: Mendelian randomisation study. *The British Journal of Psychiatry*, *205*(1), 24–28. http://doi.org/10.1192/bjp.bp.113.130419

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*(2), 481–502. http://doi.org/10.1111/j.1467-985X.2007.00527.x

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Med*, *2*(8), e124. http://doi.org/10.1371/journal.pmed.0020124

Irons, D. E., McGue, M., Iacono, W. G., & Oetting, W. S. (2007). Mendelian randomization: a novel test of the gateway hypothesis and models of gene-environment interplay. *Development and Psychopathology*, *19*(4), 1181–1195. http://doi.org/10.1017/S0954579407000612

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33 Suppl*, 245–254. http://doi.org/10.1038/ng1089

Jaffee, S. R., & Price, T. S. (2012). The implications of genotype-environment correlation for establishing causal processes in psychopathology. *Development and Psychopathology*, *24*(4), 1253–1264. http://doi.org/10.1017/S0954579412000685

Jaffee, S. R., Strait, L. B., & Odgers, C. L. (2012). From correlates to causes: Can quasi-experimental studies and statistical innovations bring us closer to identifying the causes of antisocial behavior? *Psychological Bulletin*, *138*(2), 272–295. http://doi.org/10.1037/a0026020

Jokela, M., Elovainio, M., Keltikangas-Järvinen, L., Batty, G. D., Hintsanen, M., Seppälä, I., … Kivimäki, M. (2012). Body mass index and depressive symptoms: instrumental-variables regression with genetic risk score. *Genes, Brain, and Behavior*, *11*(8), 942–948. http://doi.org/10.1111/j.1601-183X.2012.00846.x

Kendler, K. S., Neale, M. C., MacLean, C. J., Heath, A. C., Eaves, L. J., & Kessler, R. C. (1993). Smoking and major depression. A causal analysis. *Archives of General Psychiatry*, *50*(1), 36–43.

Kirkbride, J. B., Susser, E., Kundakovic, M., Kresovich, J. K., Davey Smith, G., & Relton, C. L. (2012). Prenatal nutrition, epigenetics and schizophrenia risk: can we test causal effects? *Epigenomics*, *4*(3), 303–315. http://doi.org/10.2217/epi.12.20

Kivimäki, M., Jokela, M., & Batty, G. D. (2011). Does obesity really protect against psychological distress? Examining the 'fat-jolly' versus 'fat-sad' hypotheses using Mendelian randomization. *Journal of Internal Medicine*, *269*(5), 519–520. http://doi.org/10.1111/j.1365-2796.2011.02357.x

Kivimäki, M., Jokela, M., Hamer, M., Geddes, J., Ebmeier, K., Kumari, M., … Batty, G. D. (2011). Examining overweight and obesity as risk factors for common mental disorders using fat mass and obesity-associated (FTO) genotype-instrumented analysis: The Whitehall II Study, 1985-2004. *American Journal of Epidemiology*, *173*(4), 421–429. http://doi.org/10.1093/aje/kwq444

Klengel, T., Pape, J., Binder, E. B., & Mehta, D. (2014). The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology*, *80*, 115–132. http://doi.org/10.1016/j.neuropharm.2014.01.013

Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J.-B., Rimfeld, K., von Stumm, S., … Plomin, R. (2015). Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry*. http://doi.org/10.1038/mp.2015.126

Lawlor, D. A., Harbord, R. M., Tybjaerg-Hansen, A., Palmer, T. M., Zacho, J., Benn, M., … Nordestgaard, B. G. (2011). Using genetic loci to understand the relationship between adiposity and psychological distress: a Mendelian Randomization study in the Copenhagen General Population Study of 53,221 adults. *Journal of Internal Medicine*, *269*(5), 525–537. http://doi.org/10.1111/j.1365-2796.2011.02343.x

Lewis, S. J., Araya, R., Davey Smith, G., Freathy, R., Gunnell, D., Palmer, T., & Munafò, M. (2011). Smoking is associated with, but does not cause, depressed mood in pregnancy--a mendelian randomization study. *PloS One*, *6*(7), e21689. http://doi.org/10.1371/journal.pone.0021689

Lewis, S. J., Relton, C., Zammit, S., & Davey Smith, G. (2013). Approaches for strengthening causal inference regarding prenatal risk factors for childhood behavioural and psychiatric disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(10), 1095–1108. http://doi.org/10.1111/jcpp.12127

Lutz, P.-E., & Turecki, G. (2014). DNA methylation and childhood maltreatment: from animal models to human studies. *Neuroscience*, *264*, 142–156. http://doi.org/10.1016/j.neuroscience.2013.07.069

Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., … Sullivan, P. F. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry*, *18*(4), 497–511. http://doi.org/10.1038/mp.2012.21

McClernon, F. J., & Kollins, S. H. (2008). ADHD and smoking: From genes to brain to behavior. *Annals of the New York Academy of Sciences*, *1141*, 131–147. http://doi.org/10.1196/annals.1441.016

Melchior, M., Chastang, J.-F., Head, J., Goldberg, M., Zins, M., Nabi, H., & Younès, N. (2013). Socioeconomic position predicts long-term depression trajectory: a 13-year follow-up of the GAZEL cohort study. *Molecular Psychiatry*, *18*(1), 112–121. http://doi.org/10.1038/mp.2011.116

Mill, J., & Heijmans, B. T. (2013). From promises to practical strategies in epigenetic epidemiology. *Nature Reviews. Genetics*, *14*(8), 585–594. http://doi.org/10.1038/nrg3405

Nishi, A., Numata, S., Tajima, A., Kinoshita, M., Kikuchi, K., Shimodera, S., … Ohmori, T. (2014). Meta-analyses of blood homocysteine levels for gender and genetic association studies of the MTHFR C677T polymorphism in schizophrenia. *Schizophrenia Bulletin*, *40*(5), 1154–1163. http://doi.org/10.1093/schbul/sbt154

Nitsch, D., Molokhia, M., Smeeth, L., DeStavola, B. L., Whittaker, J. C., & Leon, D. A. (2006). Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *American Journal of Epidemiology*, *163*(5), 397–403. http://doi.org/10.1093/aje/kwj062

Numata, S., Kinoshita, M., Tajima, A., Nishi, A., Imoto, I., & Ohmori, T. (2015). Evaluation of an association between plasma total homocysteine and schizophrenia by a Mendelian randomization analysis. *BMC Medical Genetics*, *16*, 54. http://doi.org/10.1186/s12881-015-0197-7

Pingault, J.-B., Côté, S., Galéra, C., Genolini, C., Falissard, B., Vitaro, F., & Tremblay, R. E. (2013). Childhood trajectories of inattention, hyperactivity and oppositional behaviors and prediction of substance abuse/dependence: a 15-year longitudinal population-based study. *Molecular Psychiatry*, *18*(7), 806–812. http://doi.org/10.1038/mp.2012.87

Relton, C. L., & Davey Smith, G. (2012a). Is epidemiology ready for epigenetics? *International Journal of Epidemiology*, *41*(1), 5–9. http://doi.org/10.1093/ije/dys006

Relton, C. L., & Davey Smith, G. (2012b). Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *International Journal of Epidemiology*, *41*(1), 161–176. http://doi.org/10.1093/ije/dyr233

Richmond, R. C., Al-Amin, A., Davey Smith, G., & Relton, C. L. (2014). Approaches for drawing causal inferences from epidemiological birth cohorts: a review. *Early Human Development*, *90*(11), 769–780. http://doi.org/10.1016/j.earlhumdev.2014.08.023

Rutter, M. (2007). *Identifying the environmental causes of disease: how should we decide what to believe and when to take action?* The Academy of Medical Sciences.

Rutter, M., Pickles, A., Murray, R., & Eaves, L. (2001). Testing hypotheses on specific environmental causal effects on behavior. *Psychological Bulletin*, *127*(3), 291–324.

Sallis, H., Steer, C., Paternoster, L., Davey Smith, G., & Evans, J. (2014). Perinatal depression and omega-3 fatty acids: a Mendelian randomisation study. *Journal of Affective Disorders*, *166*, 124–131. http://doi.org/10.1016/j.jad.2014.04.077

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. http://doi.org/10.1038/nature13595

Sheehan, N. A., Didelez, V., Burton, P. R., & Tobin, M. D. (2008). Mendelian randomisation and causal inference in observational epidemiology. *PLoS Medicine*, *5*(8), e177. http://doi.org/10.1371/journal.pmed.0050177

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. http://doi.org/10.1214/09-STS313

Stuart, E. A., & Green, K. M. (2008). Using full matching to estimate causal effects in nonexperimental studies: Examining the relationship between adolescent marijuana use and adult outcomes. *Developmental Psychology*, *44*(2), 395–406. http://doi.org/10.1037/0012-1649.44.2.395

Taylor, A. E., Fluharty, M. E., Bjørngaard, J. H., Gabrielsen, M. E., Skorpen, F., Marioni, R. E., … Munafò, M. R. (2014). Investigating the possible causal association of smoking with depression and anxiety using Mendelian randomisation meta-analysis: the CARTA consortium. *BMJ Open*, *4*(10), e006141. http://doi.org/10.1136/bmjopen-2014-006141

Taylor, A. E., Howe, L. D., Heron, J. E., Ware, J. J., Hickman, M., & Munafò, M. R. (2014). Maternal smoking during pregnancy and offspring smoking initiation: assessing the role of intrauterine exposure. *Addiction (Abingdon, England)*, *109*(6), 1013–1021. http://doi.org/10.1111/add.12514

Tobi, E. W., Goeman, J. J., Monajemi, R., Gu, H., Putter, H., Zhang, Y., … Heijmans, B. T. (2014). DNA methylation signatures link prenatal famine exposure to growth and metabolism. *Nature Communications*, *5*, 5592. http://doi.org/10.1038/ncomms6592

Walter, S., Kubzansky, L. D., Koenen, K. C., Liang, L., Tchetgen Tchetgen, E. J., Cornelis, M. C., … Glymour, M. M. (2015). Revisiting mendelian randomization studies of the effect of body mass index on depression. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *168B*(2), 108–115. http://doi.org/10.1002/ajmg.b.32286

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., … Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*, *382*(9904), 1575–1586. http://doi.org/10.1016/S0140-6736(13)61611-6

Wium-Andersen, M. K., Ørsted, D. D., & Nordestgaard, B. G. (2014a). Elevated C-reactive protein, depression, somatic diseases, and all-cause mortality: a mendelian randomization study. *Biological Psychiatry*, *76*(3), 249–257. http://doi.org/10.1016/j.biopsych.2013.10.009

Wium-Andersen, M. K., Ørsted, D. D., & Nordestgaard, B. G. (2014b). Elevated C-reactive protein associated with late- and very-late-onset schizophrenia in the general population: a prospective study. *Schizophrenia Bulletin*, *40*(5), 1117–1127. http://doi.org/10.1093/schbul/sbt120

Wium-Andersen, M. K., Ørsted, D. D., & Nordestgaard, B. G. (2015). Tobacco smoking is causally associated with antipsychotic medication use and schizophrenia, but not with antidepressant medication use or depression. *International Journal of Epidemiology*, *44*(2), 566–577. http://doi.org/10.1093/ije/dyv090

Wium-Andersen, M. K., Ørsted, D. D., Tolstrup, J. S., & Nordestgaard, B. G. (2015). Increased alcohol consumption as a cause of alcoholism, without similar evidence for depression: a Mendelian randomization study. *International Journal of Epidemiology*, *44*(2), 526–539. http://doi.org/10.1093/ije/dyu220

Wockner, L. F., Noble, E. P., Lawford, B. R., Young, R. M., Morris, C. P., Whitehall, V. L. J., & Voisey, J. (2014). Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Translational Psychiatry*, *4*, e339. http://doi.org/10.1038/tp.2013.111

Table 1: Characteristics of the studies included in the systematic review

Reference	Sample	N	Exposure	Outcome	Genetic instrument	Findings
(Lewis et al., 2011)	ALSPAC, UK	6294	Smoking	Antenatal depression	Nicotine acetylcholine receptor gene cluster (CHRNA5-A3-B4), 1 SNP	Smoking does not cause increased depression. Conversely, findings are consistent with a self-medication hypothesis, whereby depressed women smoke to alleviate their symptoms.
(Bjørngaard et al., 2013)	Norwegian HUNT study	53601	Smoking	Anxiety and Depression	Nicotine acetylcholine receptor gene cluster (CHRNA5-A3-B4), 1 SNP	Smoking is not a cause of anxiety or depression
(Taylor, Fluharty, et al., 2014)	Carta Consortium	127632	Smoking	Depression, anxiety and psychological distress	Nicotine acetylcholine receptor gene cluster (CHRNA5-A3-B4), 2 SNPs	No evidence for a causal role of smoking heaviness in the development of depression or anxiety
(Wium-Andersen, Ørsted, & Nordestgaard, 2015)	Copenhagen General Population Study and Copenhagen City Heart Study	63296	Smoking	Depression & Schizophrenia and respective medications	Nicotine acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4), 1 SNP	Smoking does not cause Depression but some evidence that it does for Schizophrenia
(Almeida et al., 2014)	Health In Men Study, Australia	3873	Alcohol	Depression	Alcohol dehydrogenase 1B (ADH1B), 1 SNP	Alcohol does not cause depression
(Wium-Andersen, Ørsted, Tolstrup, et al., 2015)	Copenhagen General Population Study and Copenhagen City Heart Study	78154	Alcohol	Depression, psychological distress, and alcoholism	Alcohol dehydrogenase 1B (ADH1B) and ADH1C, 2 SNPs	Alcohol does not cause depression but is causally linked with hospitalization/death with alcoholism

(Hung et al., 2014)	RADIANT	3222	BMI	Major depression	Fat mass and obesity-associated (FTO) gene, 1 SNP, and genetic risk score based on 32 SNPs	No evidence for a causal relationship between BMI and major depression
(Kivimäki, Jokela, Hamer, et al., 2011)	Whitehall II study	4145	BMI	Common mental disorder (anxiety and depression)	Fat mass and obesity-associated (FTO) gene, 1 SNP	MR analysis shows that BMI increases the risk of common mental disorder in men only
(Jokela et al., 2012)	Young Finns	1731	BMI	Depression	Genetic risk score, 31 SNPs	MR analysis supports a causal association between excessive BMI and increased risk of depressive symptoms
(Lawlor et al., 2011)	Copenhagen General Population Study and Copenhagen City Heart Study	53221	BMI	Psychological distress	Fat mass and obesity-associated (FTO) gene, 1 SNP and Melanocortin receptor 4 (MC4R), 1 SNP	MR analysis shows that higher BMI and WHR is associated with less psychological distress.
(Walter et al., 2015)	Female Nurse's Health Study	6989	BMI	Depression	Fat mass and obesity-associated (FTO) gene, 1 SNP and melanocortin receptor 4 (MC4R), 1 SNP, as well as a polygenic risk score based on 32 SNPs	No evidence for a causal role of BMI on depression
(Sallis et al., 2014)	ALSPAC, UK	3397	Docosahexaenoicacid (DHA) and eicosapentaenoicacid (EPA)	Perinatal, antenatal and postnatal depression	Polygenic risk scores, EPA, 23 SNPs and DHA, 4 SNPs	No evidence that fatty acids are causally related to depression outcomes

(Almeida et al., 2009)	Health In Men Study, Australia	3700	C-reactive protein (CRP)	Depression	CRP gene, 2 SNPs	Depressive symptoms in later life are unlikely caused by an increase in the plasma level of CRP
(Wium-Andersen, Ørsted, & Nordestgaard, 2014a)	Copenhagen General Population Study and Copenhagen City HearStudy	78,809	C-reactive protein (CRP)	Depression, psychological distress and major somatic diseases	CRP gene, 4 SNPs	No evidence that increased C-reactive protein levels cause depression
(Wium-Andersen, Ørsted, & Nordestgaard, 2014b)	Copenhagen General Population Study and Copenhagen City HearStudy	78810	C-reactive protein (CRP)	Schizophrenia	CRP gene, 4 SNPs	No significant evidence that elevated C-reactive protein is causally related to schizophrenia, although a causal relationship cannot be excluded
(Nishi et al., 2014)	Meta-analysis of 6 Japanese studies	10,378	homocysteine	Schizophrenia	Methylenetetrahydrofolate reductase (MTHFR), 1 SNP	Evidence of causal relationship between higher homocysteine levels and higher risk of schizophrenia
(Numata et al., 2015)	Meta-analysis of 36 case control studies	25,599	homocysteine	Schizophrenia	Methylenetetrahydrofolate reductase (MTHFR), 1 SNP	Evidence of causal relationship from higher homocysteine levels to higher risk of schizophrenia
(Irons et al., 2007)	Sibling Interaction and behaviour Study	180	Alcohol	Nicotine dependence, drug abuse and dependence, antisocial personality disorder	Aldehyde dehydrogenase 2 (ALDH2), 1 SNP	Evidence of causal relationship from higher homocysteine levels to higher risk of schizophrenia
(Taylor,	ALSPAC, UK	1020	Maternal	Offspring smoking	Nicotine acetylcholine	No evidence that maternal

Howe, et al., 2014)	smoking	initiation	receptor gene cluster (CHRNA5-A3-B4), 1 SNP	smoking is causally related to offspring smoking initiation in adolescence

Figure 1: Representation of Mendelian randomisation

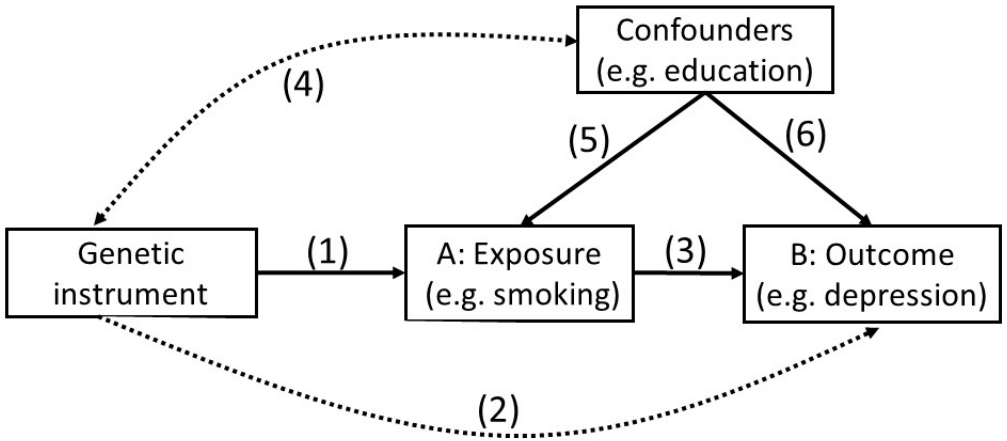

Figure Legend: Mendelian randomisation uses a genetic instrument associated with an exposure in order to test whether the association between the exposure A and an outcome B (path 3) is causal. An adequate genetic instrument must be robustly associated with the exposure (solid path 1), but not be associated with the confounders (dashed path 4). The latter condition is key so that MR estimation is not affected by

variables that confound the association between the exposure and the outcome in conventional observational research (5 & 6). Following the principles of instrumental variables, a significant association between the genetic instrument and the outcome is interpreted as evidence of the causal relationship between the exposure and the outcome. Importantly, all the effect of the genetic instrument on the outcome should happen through the exposure. Therefore, no direct effect must remain once the exposure is taken into account. This is why path (2) is dashed in the Figure, which includes the exposure. However, it is important to note that the observed value of path (2) is expected to be significant, which is how we assess if there is a causal relationship between A and B. This is similar to a full mediation analysis where the direct path between a predictor X and an outcome Y (path often labelled c) is not significant when the effect of the mediator M is taken into account (i.e. the new path, labelled c', is not significant). Finally, the causal estimate of path (3) can be estimated based on the observed values of path (1) and (2).

Figure 2: Mendelian randomisation and randomised controlled trial